#### CSE574 Introduction to Machine Learning

ternative Vi

Regression

Support Vector Machine

Large Margin

The Mathematics behind Large Margin

# CSE574 Introduction to Machine Learning Support Vector Machine

Jue Guo

University at Buffalo

February 4, 2024

## Outline

CSE574 ntroduction to Machine Learning

Jue Gi

Alternative Vie of Logistic Regression

Support Vector Machine

Large Margin Intuition

behind Large Margin Classification

Kernel:

1 Alternative View of Logistic Regression

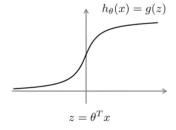
- 2 Support Vector Machine
  - Large Margin Intuition
  - The Mathematics behind Large Margin Classification

# Alternative View of Logistic Regression

Alternative View of Logistic Regression

A quick review:  $h_{\theta}(x) = \frac{1}{1+e^{-\theta^T x}}$ 

- if y = 1, we want  $h_{\theta}(x) \approx 1$ ,  $\theta^T x \gg 0$
- if y = 0, we want  $h_{\theta}(x) \approx 0$ ,
  - $\theta^T x \ll 0$



The cost of a single example:

$$- (y \log h_{\theta}(x) + (1 - y) \log (1 - h_{\theta}(x)))$$

$$= -y \log \frac{1}{1 + e^{-\theta^{T}x}} - (1 - y) \log \left(1 - \frac{1}{1 + e^{-\theta^{T}x}}\right)$$

#### CSE574 ntroduction to Machine Learning

Jue Gu

Alternative View of Logistic Regression

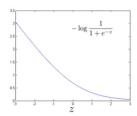
Support Vector

Intuition

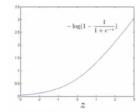
behind Large Margin Classification

$$-y \log \frac{1}{1 + e^{-\theta^T x}} - (1 - y) \log \left(1 - \frac{1}{1 + e^{-\theta^T x}}\right)$$

if 
$$y = 1$$
 (want  $\theta^T x \gg 0$ )



if 
$$y = 0$$
 (want  $\theta^T x \ll 0$ )



# Support Vector Machine

CSE574 ntroduction to Machine Learning

Jue Gu

Alternative View of Logistic Regression

Support Vector Machine

Large Margir

behind Large Margin

Kernel

## **Cost Function of Logistic Regression**

$$\min_{\theta} \frac{1}{m} \left[ \sum_{i=1}^{m} y^{(i)} \left( -\log h_{\theta} \left( x^{(i)} \right) \right) + \left( 1 - y^{(i)} \right) \left( -\log \left( 1 - h_{\theta} \left( x^{(i)} \right) \right) \right) \right] + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_{j}^{2}$$

#### **Cost Function of Support Vector Machine**

$$\min_{\theta} C \sum_{i=1}^{m} \left[ y^{(i)} \operatorname{cost}_{1} \left( \theta^{T} x^{(i)} \right) + \left( 1 - y^{(i)} \right) \operatorname{cost}_{0} \left( \theta^{T} x^{(i)} \right) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_{j}^{2}$$

# Large Margin Intuition

CSE574 ntroduction to Machine Learning

Alternative View of Logistic Regression

Support Vecto Machine

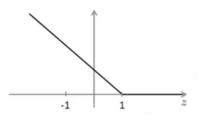
Large Margin

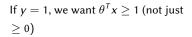
The Mathematics behind Large Margin

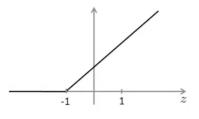
Vornole

## **Support Vector Machine**

$$\min_{\theta} C \sum_{i=1}^{m} \left[ y^{(i)} \operatorname{cost}_{1} \left( \theta^{T} x^{(i)} \right) + \left( 1 - y^{(i)} \right) \operatorname{cost}_{0} \left( \theta^{T} x^{(i)} \right) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_{j}^{2}$$







If 
$$y = 0$$
, we want  $\theta^T x \le -1$  (not just  $< 0$ )

Kernel

## **Support Vector Machine**

$$\min_{\theta} C \sum_{i=1}^{m} \left[ y^{(i)} \cot_{1} \left( \theta^{T} x^{(i)} \right) + \left( 1 - y^{(i)} \right) \cot_{0} \left( \theta^{T} x^{(i)} \right) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_{i}^{2}$$

Given that C is a very large value, we want that the first term to be 0. Let's try to understand the optimization problem in the context of what would it take to make this first term in the objective equal to 0.

Whenever 
$$y^{(i)} = 1, \theta^{\top} x^{(i)} \geqslant 1;$$
 Whenever  $y^{(i)} = 0, \theta^{\top} x^{(i)} \leqslant -1$ 

Now, the optimization problem can be written as:

$$\min C \cdot 0 + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2$$
s.t.  $\theta^\top x^{(i)} \geqslant 1$  if  $y^{(i)} = 1$ 

$$\theta^T x^{(i)} \leqslant -1$$
 if  $y^{(i)} = 0$ 

#### CSE574 ntroduction to Machine Learning

Jue Guo

Alternative View of Logistic Regression

Support Vecto Machine

Large Margin

The Mathematics behind Large Margin

Kernels

### SVM Decision Boundary: Linearly separable case

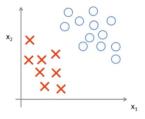


Figure: Linearly Separable Case

#### CSE574 ntroduction to Machine Learning Jue Guo

Alternative View of Logistic Regression

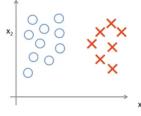
Support Vecto Machine

#### Large Marg

The Mathematics behind Large Margin

Kernels

# Large margin classifier in presence of outliers



# The Mathematics behind Large Margin Classification

CSE574
Introduction to
Machine
Learning

ternative Viev Logistic

Support Vecto

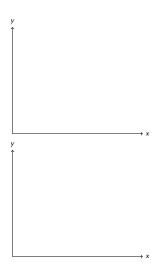
Machine

The Mathematics

behind Large Margin Classification

Kernels

## **Vector Inner Product**



$$u = \left[ \begin{array}{c} u_1 \\ u_2 \end{array} \right] \quad v = \left[ \begin{array}{c} v_1 \\ v_2 \end{array} \right]$$

# **SVM** Decision Boundary

CSE574 ntroduction to Machine Learning

Alternative Vi of Logistic Regression

Support Vecto

Large Margin

The Mathematics behind Large Margin Classification

$$\min_{\theta} \quad \frac{1}{2} \sum_{j=1}^{n} \theta_{j}^{2}$$
s.t. 
$$\theta^{T} x^{(i)} \ge 1 \quad \text{if } y^{(i)} = 1$$

$$\theta^{T} x^{(i)} \le -1 \quad \text{if } y^{(i)} = 0$$



Support Vecto Machine

Large Marg Intuition

The Mathematics behind Large Margin Classification

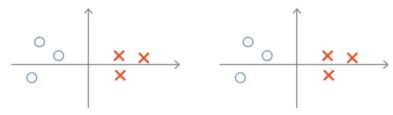
Kernels

$$\begin{aligned} & \min_{\theta} & & \frac{1}{2} \sum_{j=1}^{n} \theta_{j}^{2} = \frac{1}{2} \|\theta\|^{2} \\ & \text{s.t.} & & p^{(i)} \cdot \|\theta\| \geq 1 & \text{if } y^{(i)} = 1 \\ & & & p^{(i)} \cdot \|\theta\| < -1 & \text{if } v^{(i)} = 0 \end{aligned}$$

where  $p^{(i)}$  is the projection of  $x^{(i)}$  onto the vector  $\theta$ . Simplification:  $\theta_0 = 0$ ; this simplification merely makes the decision boundary to pass through (0,0);

## **Bad Decision Boundary**

# **Good Decision Boundary**



# **Reading Assignments**

CSE574 ntroduction to Machine Learning

Alternative Vic of Logistic Regression

Support Vecto

The Mathematics behind Large Margin Classification

- Why the parameter vector orthogonal to the decision boundary?
  - Orthogonality in Neural Network
  - SVM

# Non-linear Decision Boundary

CSE574 ntroduction to Machine Learning

Jue Gu

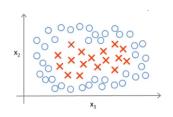
Alternative View of Logistic Regression

Support Vecto Machine

Large Margin

The Mathematic behind Large Margin

Kernels



Predict y = 1 if

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \dots \ge 0$$

$$h_0(x) = \begin{cases} 1 & \text{if } \theta_0 + \theta_1 x_1 + \cdots \geqslant 0 \\ 0 & \text{otherise.} \end{cases}$$

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \cdots$$
  
 $f_1 = x_1, \quad f_2 = x_2, \quad f_3 = x_1 x_2,$   
 $f_4 = x_1^2, \quad f_5 = x_2^2, \dots$ 

Is there a different/ better choice of the features  $f_1, f_2, f_3, \ldots$ ?

## Kernel

CSE574 introduction to Machine Learning

Jue Gui

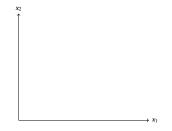
Alternative View of Logistic Regression

Support Vector Machine

Large Margir

behind Large
Margin

Kernels



Given x, compute new feature depending on proximity to landmarks  $l^{(1)}$ ,  $l^{(2)}$ ,  $l^{(3)}$ ;

#### Given x:

• 
$$f_1 = \text{similarity}(x, l^{(1)}) = \exp(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2})$$

• 
$$f_2 = \text{similarity}(x, l^{(2)}) = \exp(-\frac{\|x - l^{(2)}\|^2}{2\sigma^2})$$

$$f_3 = similarity(x, l^{(3)}) = exp(...)$$

#### CSE574 ntroduction to Machine Learning

Jue Guo

Alternative Viev of Logistic Regression

Support Vecto

Large Margin

behind Large Margin

Classification Kernels Questions?